HASSELBLAD et al. 1998

45043

RESTRICTED VERSUS UNRESTRICTED MODELS: WHY IS ONE THE MAGIC NUMBER? V Hasselblad[1], A M Jarabek[2], D J Svendsgaard[3], and J M Davis[2]. [1]Center for Health Policy Research and Education, Duke University, Durham, NC 27705, [2]Environmental Criteria and Assessment Office, U.S. EPA, Research Triangle Park, NC 27711, [3]formerly with the Health Effects Research Laboratory, U.S. EPA, Research Triangle Park, NC 27711.

ABSTRACT

The Weibull model has been proposed for calculation of a benchmark dose (BMD) for dose-response assessment of noncancer toxicity. As defined by Crump (1984), the model is

$$P(response) = \gamma + (1 - \gamma)(1 - \exp(-\beta x^{\alpha})),$$

where P is probability, x is the exposure concentration, and α, β, and γ are parameters to be estimated (γ is the non-zero background rate of response). The parameter α has been restricted to be ≥ 1 by some authors, as has the slope of the log-logistic regression model. The restriction of $\alpha \geq 1$ guarantees that the slope is not infinite at the origin, and eliminates the possibility of fitting a supralinear model to the data. However, the slope is finite even in the unrestricted case when plotted on a logarithmic scale. Restriction of $\alpha \geq 1$ has some historical precedent in approaches to cancer risk assessment but is basically an arbitrary constraint and can result in significantly poor estimation of the BMD. Supralinear models (with $\alpha < 1$) may better fit data that reflect saturation phenomena or the responses of a sensitive subpopulation. Restricted and unrestricted models (Weibull and log-logistic) are applied to an epidemiologic data set to illustrate the impact on resultant BMD estimates and goodness-of-fit.

The 34th Annual Meeting of the Society of Toxicology
Baltimore, MD, March 5-9, 1995

The Toxicologist 15(1): 178 (Abstract No. 952).

INTRODUCTION

The benchmark approach has recently been proposed for dose response analysis of both noncancer (Barnes et al., 1994) and cancer toxicity (U.S. EPA, 1994). As proposed by Crump (1984), the approach involves fitting a mathematical model to the available toxicity data to obtain a maximum likelihood estimate (MLE) of the dose associated with a specified increase in toxic response (e.g., 10% for dichotomous response data). This MLE is denoted as the benchmark dose (BMD). A lower confidence limit (e.g., the 90th percentile) is then calculated for the BMD and denoted as the BMDL. The BMDL has been proposed as a substitute for a NOAEL in deriving oral reference doses or inhalation reference concentrations (Crump, 1984; Barnes et al., 1994) or for deriving unit risk estimates for cancer assessment (U.S. EPA, 1994).

One mathematical model proposed for calculation of the benchmark calculation is the Weibull. As defined by Crump (1984), the model is

$$P(response) = \gamma + (1 - \gamma)(1 - \exp(-\beta x^{\alpha})),$$

where P is probability, x is the exposure concentration, and α, β, and γ are parameters to be estimated (γ is the non-zero background rate of response). Another model used often by epidemiologists is the log-logistic model,

$$P(\text{response}) = \gamma + (1 - \gamma)\,\Phi[\alpha + \beta \log(x)], \qquad \text{(A-4)}$$

where $\Phi(x)$ is the cumulative logistic distribution, and α, β and γ are parameters to be estimated. The parameter α in the Weibull model has been restricted to be ≥ 1 by some authors, as has the slope (β) of the log-logistic regression model. The restriction of $\alpha \geq 1$ guarantees that the slope is not infinite at the origin, and eliminates the possibility of fitting a supralinear model to the data.

As explained by Crump (1984), these restrictions have their basis in dose-response modeling for cancer data:

> Note that the restrictions $\alpha \geq 1$ and $b \geq 1$ are assumed for the Weibull and log-normal models, respectively. Some restrictions of this nature seem necessary with these models; otherwise these models can exhibit very extreme and biologically implausible behavior. The restriction $\alpha \geq 1$ was selected for the Weibull model because $\alpha < 1$ corresponds to a supralinear curve shape which is implausible for any biological effect. *Although the restriction $b \geq 1$ for the slope parameter in the log-normal model does not have a strong theoretical basis, it does have precedent as it was recommended by Mantel et al. (1975) for cancer data.* [Emphasis added].

Given that the basis for restricting the slope parameter is attributed to a recommendation by Mantel et al. (1975) regarding cancer data, a more extensive consideration of the theoretical basis for cancer modeling at that time may help illuminate issues for the application of such models to benchmark analyses of current bioassays for both cancer and noncancer toxicity data.

Originating in the late 1970s to 1980s, cancer dose-response modeling used the log-probit and then the "Mantel-Bryan" model, the latter being essentially the log-probit for a one-dose estimate (Mantel and Bryan, 1961; Mantel et al., 1975). The available data at the time were largely from screening studies that generally used single dose levels. The prevailing understanding of the pathogenesis process was based on radiation carcinogenesis and the view that only one "hit" was required to initiate cancer. With the advent of multiple dose studies in the later 1980s, initiation-promotion studies, and advances in understanding of mechanisms of carcinogenesis, a multiple-stage model was needed. Armitage and Doll (1957) proposed the linearized multistage model approach in which a general polynomial (multistage) model was fit to the data and the maximum linear term consistent with the data was calculated. The low-dose slope of the dose-response function was then equated to the coefficient of this maximum linear term and the slope was equated to the upper bound of potency. A restriction on the coefficients of dose (β) ensured that the model increases monotonically with dose, whereas the power parameters on each of the dose terms (α) were usually positive integers because it was believed that a minimum of one hit could initiate cancer.

Another, unstated reason for the use of restricted models is that they eliminate certain computational difficulties. Confidence intervals for nonlinear models are usually calculated from the likelihood function, which is assumed to be approximately normal. The variances are often calculated from the second partial derivatives and the confidence intervals are calculated assuming normality. Approximate lower confidence limits are difficult to compute when the slope is very steep near the origin. This difficulty does not imply that the model is incorrect but that the statistical methods have difficulty with the data. If a logarithmic scale is used for the x-axis, a common approach, the slope is finite everywhere for all α and β, and the standard S-shaped sigmoidal curve results. Thus, there is no reason *a priori* to rule out the use of the unrestricted models. A solution to this computational problem would be to use a Bayesian approach that yields exact results (subject to numerical accuracy).

Because a mathematical model is a tool to describe the observed dose-response behavior, the impact of the restriction of the power parameter on dose is explored using an epidemiologic data set. The appropriateness of this mathematical convenience taking priority over fitting the observed data is discussed.

METHODS

Health effects data. The epidemiologic data of Roels et al. (1992) were used for the analysis. Roels et al. (1992) measured neurobehavioral function (eye-hand coordination, visual reaction time, and hand steadiness) in 101 matched control "nonexposed" male workers (actually, exposed to unmeasured but presumably very low levels of inhaled Mn) and 92 alkaline battery plant male workers who were exposed to levels ranging from 40 to 4,433 μg Mn/m^3. Eye-hand coordination (EHC) was chosen as the health endpoint for dose-response analysis for this presentation because it appeared to be the most sensitive indicator. Because the various neurobehavioral tests were not strictly independent, only one representative parameter per test was selected. Roels et al. (1992) designated EHC "abnormal" when the percent precision value was lower than the 5th percentile of the control group. In this way the continuous measure of EHC was "dichotomized" into a quantal variable. In a personal communication, Roels (1993) provided individual data for exposure with corresponding EHC responses as "abnormal" or "normal". These individual data were used to fit the mathematical models.

Exposure measure. Roels et al. (1992) and Roels (1993) described two measures of respirable Mn dust: 1) the lifetime integrated respirable dust concentration (LIRD), expressed as μg Mn/m^3 x years, and 2) the current concentration of respirable dust (CRD), expressed as μg Mn/m^3. The CRD was measured at the time the study was conducted and refers to a representative concentration measured for the type of job performed by a worker (e.g., electrician, maintenance worker). The LIRD value for each worker was constructed from the CRD value by summing over the worker's entire period of employment. If a worker changed jobs within the plant during his period of employment, the CRD for each job held was multiplied by the number of years the worker performed that job. The LIRD was calculated by summing across jobs. Because Mn and/or its toxicity has the potential to accumulate, the LIRD was chosen as the exposure measure.

Mathematical modeling. The Weibull model can be written as

$$P\ (\text{response}) = \gamma + (1 - \gamma)\ [1 - \exp(-\beta x^{\alpha})], \qquad \text{(A-1)}$$

where x is the exposure concentration and α, β, and γ are parameters to be estimated. In particular, γ is the nonzero background rate of response. The restricted Weibull model is where α is restricted to be ≥ 1. The unrestricted model does not set constraints on this parameter, but rather allows the best fit to the data.

Another model used often by epidemiologists is the log-logistic model,

$$P(\text{response}) = \gamma + (1 - \gamma)\,\Phi\,[\alpha + \beta \log(x)], \qquad \text{(A-4)}$$

where $\Phi(x)$ is the cumulative logistic distribution, and α, β and γ are parameters to be estimated. As was done with the Weibull model, β was either restricted to be ≥ 1 or was allowed to be estimated by the data.

Benchmark dose. From the models applied to the experimental data, the benchmark dose (BMD) is defined as the MLE for a specified health effect. The specified health effect defined for the purposes of this presentation was a 10% increase in abnormal EHC. The lower 90% confidence limit on the BMD (BMDL) was also calculated.

Bayesian approach. As described elsewhere (see Jarabek and Hasselblad poster #950 this session) a Bayesian approach can be used to analyze dose-response data. The approach is the same as the benchmark approach except that it expresses the BMD as a probability density (posterior distribution). The BMDL is analogous to the 5th percentile of this distribution. Because the approach uses numeric integration to calculate the parameter estimates, the problem with approximation of the confidence interval is obviated. Also, the graphical display of skewness (nonnormality) can aid the decision of whether the data provide meaningful information to dose-response assessment. Computation of a confidence profile is described briefly here. For a more detailed description, refer to Eddy et al. (1992) and Hasselblad and Jarabek (1995).

The basic formula of Bayesian statistics is:

$$p'(\beta) = L(\beta \mid \text{data})\, p(\beta),$$

where:

β = parameter of interest,
$p(\beta)$ = prior distribution for β,
$L(\beta \mid \text{data})$ = likelihood for β given the new data, and
$p'(\beta)$ = posterior distribution for β.

Since $p'(\beta)$ will become the prior for the next experiment, it is denoted by the same letter.

Because basic parameters are not functions of other parameters in the model, they must have priors defined. We assume that there is no positive benefit from the toxic chemical and restrict the prior on β to be positive. The choice of a prior on the background rate has minimal effect. The parameter of interest, the dose producing a specified health effect, is a functional parameter and is determined from the priors and the likelihood function. Figure 1 illustrates the Bayesian approach. A bioassay was conducted to determine information about β, resulting in the likelihood, $L(\beta)$, shown as the dotted line. The horizontal dashed line represents the prior for β based on the assumption that there no positive benefit of exposure (i.e., $p(\beta) = 1$ for $\beta > 0$, otherwise $p(\beta) = 0$). The posterior distribution, $p'(\beta)$, is the product of the two prior distributions, properly normalized to be a probability distribution, and is shown as a solid line in Figure 1. Note that this distribution has the same general shape as the likelihood function, except that it has no mass below zero (a truncated distribution). The posterior distribution of d can be calculated from the posterior distribution of β. This posterior distribution represents the probability that this dose could cause the specified health effect. The posterior distribution, $p'(\beta)$, can be used as a prior if another experiment is conducted giving additional information about d and β, and the application of Bayes' formula repeated.

RESULTS

Figure 2 shows the graph of the restricted Weibull model fit to the EHC data. Table 1 provides the goodness of fit statistics. The restricted log-logistic model (graphical display not shown) fit the data very similarly. The log-likelihood (a measure of goodness of fit) was -71.0014 for the log-logistic model.

Figure 3 shows the graph of the unrestricted Weibull model fit to the EHC data. Table 2 provides the goodness of fit statistics. The unrestricted log-logistic model (graphical display not shown) fit the data very similarly. The log-likelihood (a measure of goodness of fit) was -69.1790 for the log-logistic model.

Figure 4 shows the confidence profiles for the restricted and unrestricted Weibull models calculated using the Bayesian approach.

Table 3 shows the BMD and BMDL values for EHC for each model using both the benchmark and Bayesian approaches.

DISCUSSION / CONCLUSIONS

As shown in Table 3, use of restricted models allows the calculation of BMD and BMDL values from these data. Close examination of Figures 2 and 3, however, indicate why the calculation of these values may be deceiving. As shown in Figure 2, use of the restricted model results in an MLE that overestimates virtually all of the response data at the high concentrations. Figure 3 shows a much better fit to all of the data. The goodness of fit statistics shown in Tables 2 and 3 support this assertion (e.g., the log-likelihood for restricted versus unrestricted Weibull models is -69.1765 and -71.2001, respectively).

Use of the restricted models is in part motivated by computational difficulties. As noted in Table 3, lower confidence limits could not be calculated for the benchmark approach for either the unrestricted Weibull or log-logistic models. The Bayesian approach does not have these problems since it uses numeric integration to calculate the parameter estimates. The BMD and BMDL values are shown to be very low for the unrestricted models. The posterior distributions for the unrestricted model shown in Figure 4 shows that there is great uncertainty. This results from a wide range of possible parameter values which give similar fits. The restricted model creates an illusion of precision that is not justified by the data.

As noted in the introduction, the use of restricted models has also been based on the assertion that the restriction prevents "implausible behavior". However, even with the restrictions as defined, the restricted models do not rule out implausible fits. Figure 5 illustrates an example. Panel a of Figure 5 shows a simple data set (artificial) that is almost perfectly fit by a restricted Weibull model. Panel 5b shows an unrestricted Weibull model fit to the same data, and would be deemed "implausible". Panel 5c, however, shows another fit that is within the confidence region of the restricted model. The fit in panel 5c could also be argued to be implausible, but it is allowed with the restricted model. Thus, the restricted models do not ensure that implausible fits are ruled out.

The flaw in promoting only the use of restricted models is that it suggests that a general mathematical convenience should take priority over fitting the observed dose-response data. In this example, the epidemiologic data chosen were fit better by an unrestricted model (either Weibull or log-logistic). Once the fit is established the biologic motivation for such should be explored. It is possible that the supralinear dose-response relationship described by the unrestricted models is in fact an appropriate reflection of the presence of "sensitive" workers in the Roels et al. (1992) cohort. Thus, at least considering the unrestricted models is important, since a subpopulation of susceptible persons is certainly not "biologically implausible".

This presentation suggests that the adoption of the benchmark approach to endpoints other than dichotomous developmental data in laboratory animals should be done with caution and without the constraints superimposed by the use of restricted models. This would allow for development of a better understanding of how other endpoints (e.g., neurobehavioral function) can be best described by mathematical models for purposes of dose-response assessment.

Table 1. Results of fitting the restricted Weibull model to the data of Roels et al. (1992)

Variable	Beta	Std. Error Beta
Slope (β)	1.3290E-04	5.0116E-05
Power (α)	1	0.52574
Background (γ)	7.1633E-02	2.5493E-02

Goodness of fit

Log-likelihood (fitted model): -71.2001

Concentration[1]	Expected	Observed	Size
0.0	7.23	5	101
62.0	0.63	2	8
236.0	0.40	0	4
427.0	1.23	0	10
587.0	2.26	7	16
763.0	1.29	3	8
1009.0	1.13	1	6
1132.0	0.81	0	4
1372.0	2.04	0	9
1656.0	1.53	3	6
1921.0	1.97	0	7
2541.0	2.36	2	7
3037.0	0.76	0	2
3619.0	1.70	3	4
4433.0	0.48	0	1

Chi-squared = 32.31 for 12 d.f., p-value = 0.0012.

[1] Individual exposure concentration data were grouped into intervals that would minimize the variation within the group in order to calculate goodness of fit statistics.

Table 2. Results of fitting the unrestricted Weibull model to the data of Roels et al. (1992)

Variable	Beta	Std. Error Beta
Slope (β)	7.6103E-02	0.1429
Power (α)	0.1493	0.2700
Background (γ)	4.9565E-02	2.1608E-02

Goodness of fit

Log-likelihood (fitted model): -69.1765

Concentration[1]	Expected	Observed	Size
0.0	5.01	5	101
62.0	1.40	2	8
236.0	0.80	0	4
427.0	2.12	0	10
587.0	3.51	7	16
763.0	1.81	3	8
1009.0	1.39	1	6
1132.0	0.94	0	4
1372.0	2.16	0	9
1656.0	1.47	3	6
1921.0	1.74	0	7
2541.0	1.79	2	7
3037.0	0.52	0	2
3619.0	1.06	3	4
4433.0	0.27	0	1

Chi-squared = 24.02 for 12 d.f., p-value = 0.0202.

[1] Individual exposure concentration data were grouped into intervals that would minimize the variation within the group in order to calculate goodness of fit statistics.

Table 3. BMD and BMDL Values (LIRD in $\mu g\ Mn/m^3$ x years) for a 10% Increase in Abnormal Eye-Hand Coordination (EHC) Response by Different Models

Benchmark approaches	BMD	BMDL
Weibull restricted ($\alpha \geq 1$)	793	465
Weibull unrestricted	9	(*)[1]
Log-logistic restricted ($\beta \geq 1$)	683	370
Log-logistic unrestricted	9	(*)
Bayesian approaches		
Weibull restricted ($\alpha \geq 1$)	864	508
Weibull unrestricted	16	0.06
Log-logistic restricted ($\beta \geq 1$)	1,322	601
Log-logistic unrestricted	331	1

[1] Numerically intractable. Computational difficulties resulted from exponentiating values with large absolute magnitudes.

REFERENCES

Armitage, P. and R. Doll. 1975. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br. J. Cancer 11: 161-169.*

Barnes, D.G., G.P. Datson, J.S. Evans, A.M. Jarabek, R.J. Kavlock, C.A. Kimmel, C. Park, and H.L. Spitzer. 1994. Benchmark dose workshop: Criteria for Use of a benchmark dose to estimate a reference dose. *Regul. Toxicol. Pharmacol. 20: In press.*

Crump, K.S. 1984. A new method for determining allowable daily intakes. *Fund. Appl. Toxicol. 4: 845-871.*

Eddy, D.M., V. Hasselblad, and R.D. Shachter. 1992. The statistical synthesis of evidence: Meta analysis by the Confidence Profile Method, Academic Press.

Hasselblad, V. and A.M. Jarabek. 1995. Dose-response analysis of toxic chemicals. In: Applied Bayesian methods for biostatisticians, (Eds. D.A. Berry and D.K. Stangl), Marcel Dekker, NY. In press.

Mantel, N., N.R. Bohidar, C.C. Brown, J.L. Ciminera, and J.W. Tukey. 1975. An improved Mantel-Bryan procedure for "safety" testing of carcinogens. *Cancer Res. 35: 865-872.*

Mantel, N. and W.R. Bryan. 1961. "Safety" testing of carcinogenic agents. *J. Natl. Cancer Inst. 27: 455-470.*

U.S. Environmental Protection Agency. 1994. Draft revisions to the guidelines for carcinogen risk assessment. Office of Research and Development, Office of Health and Environmental Assessment, EPA report no. EPA/600/BP-92/003 August 1994 External Review Draft. Washington D.C.

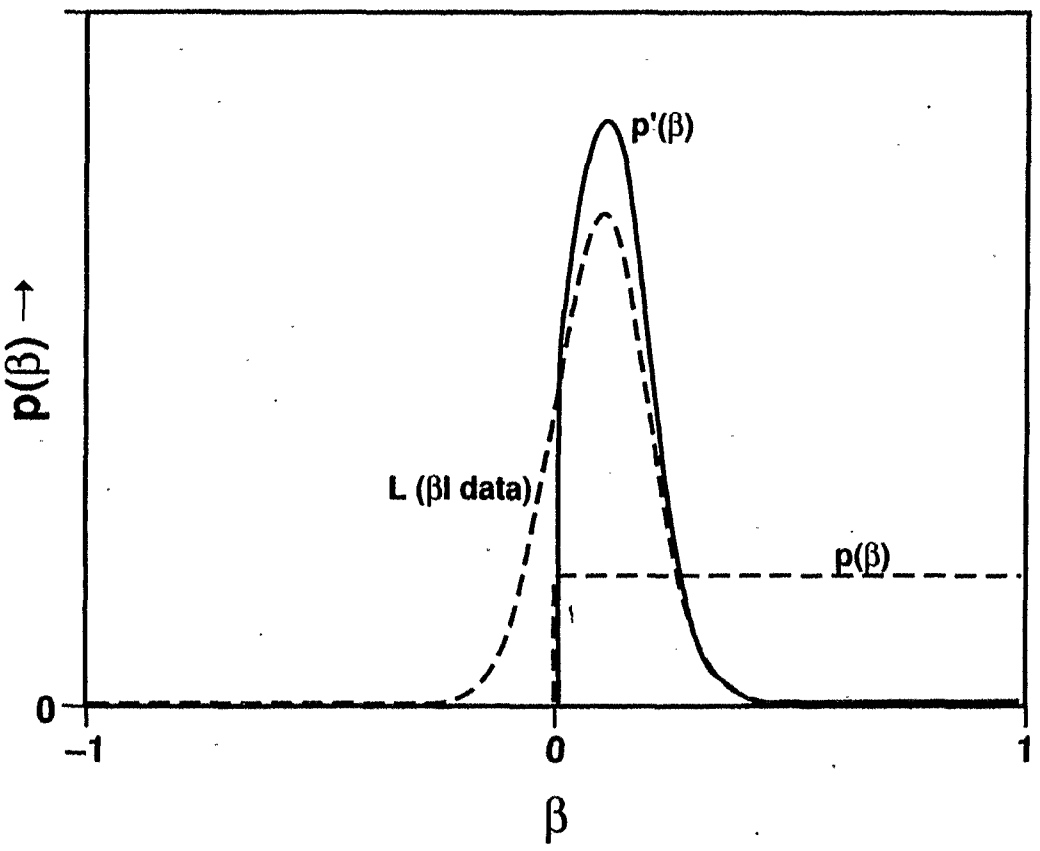

Figure 1. Schematic of computing a posterior distribution [p'(β)] from a likelihood function [L(βl data)] and a prior distribution [p(β)].

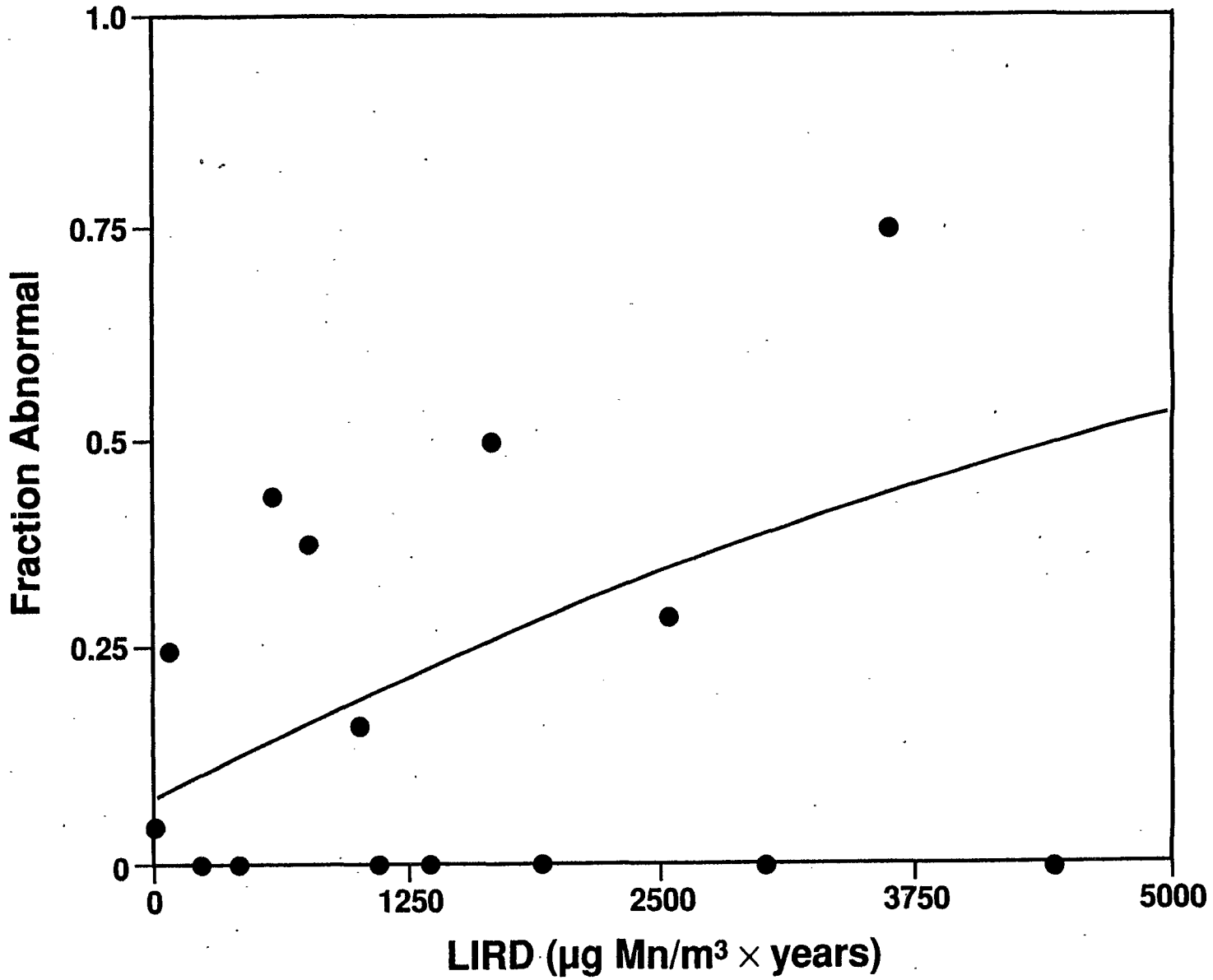

Figure 2. Fit of the Restricted Weibull model to the EHC data.

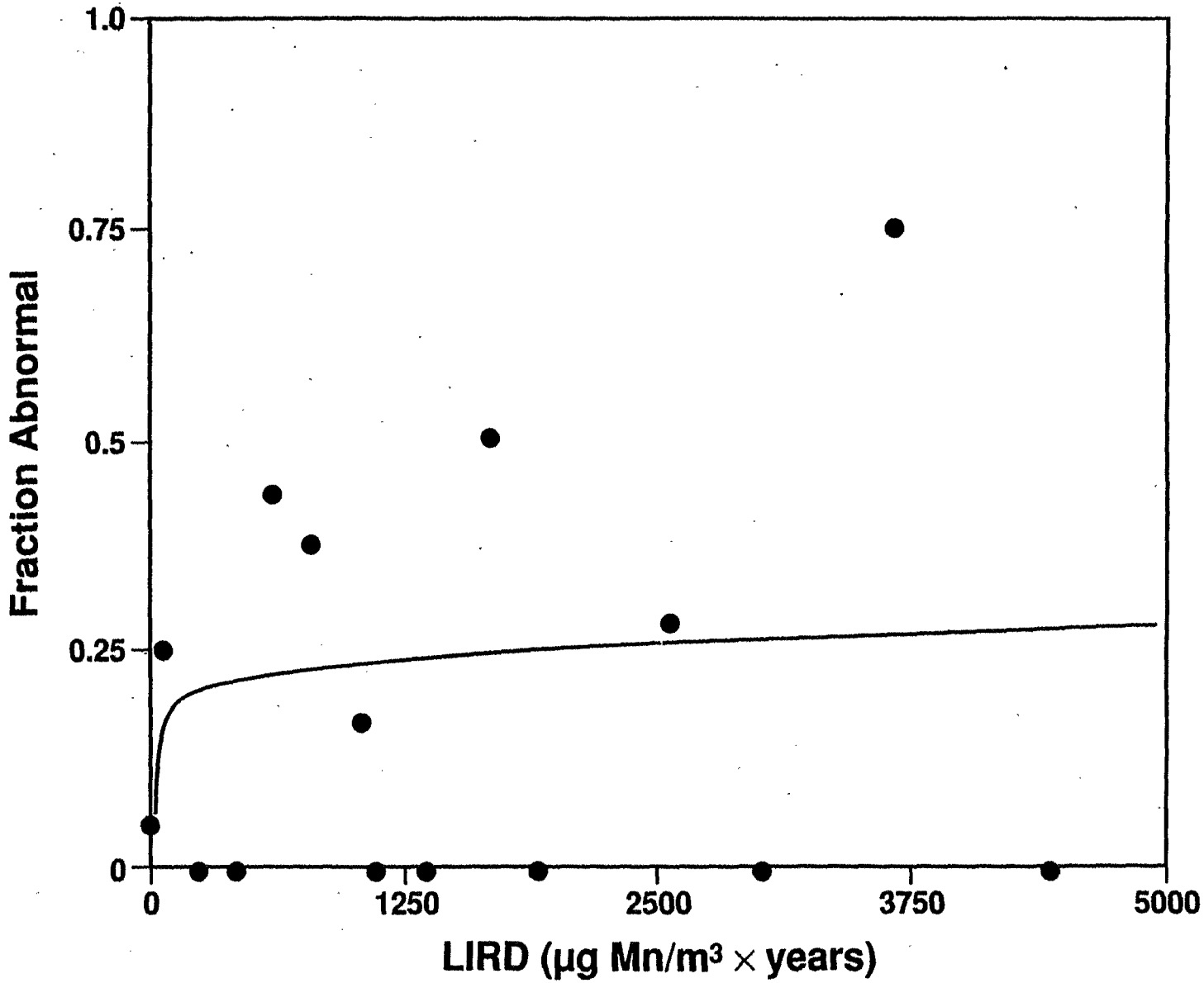

Figure 3. Fit of the Unrestricted Weibull model to the EHC data.

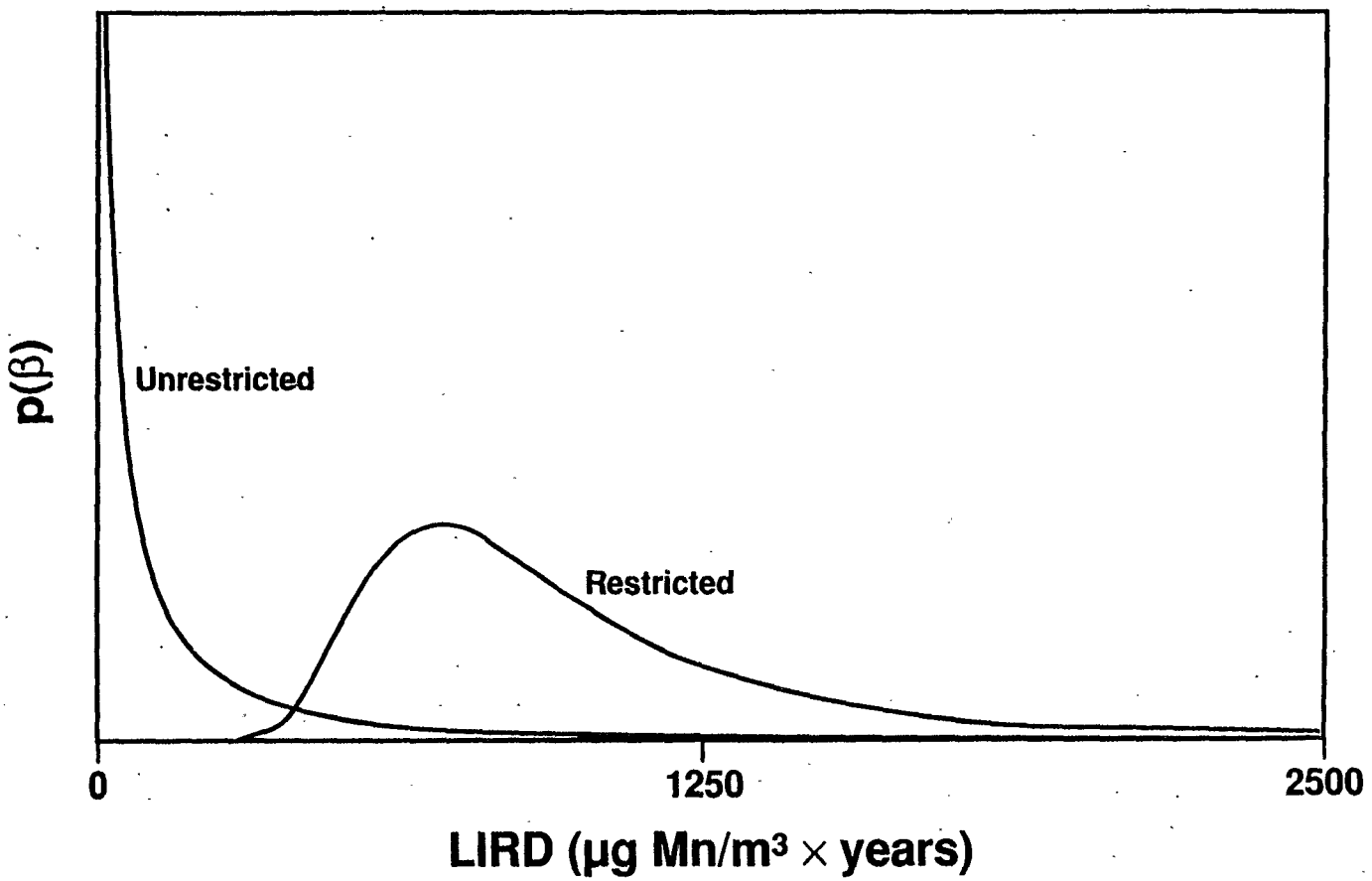

Figure 4. Posterior distributions for the restricted and unrestricted Weibull models calculated using the Bayesian approach.

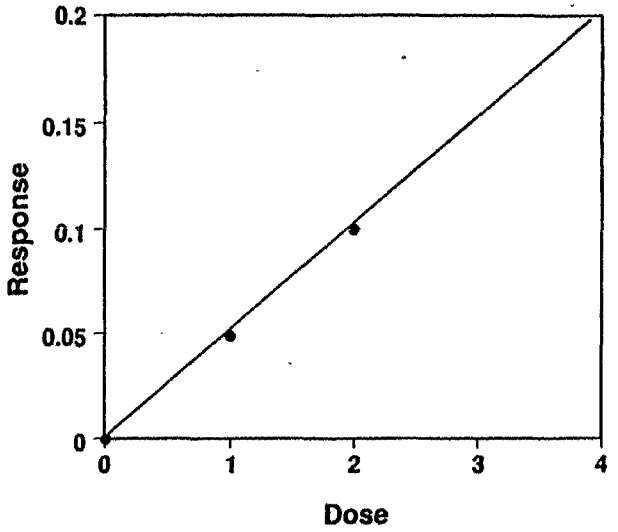

Figure 5a. Optimal fit of restricted Weibull model to artificial data.

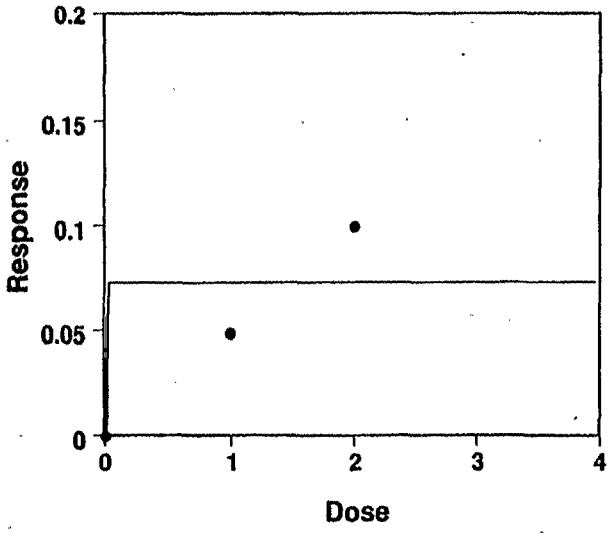

Figure 5b. "Implausible" fit of unrestricted Weibull model to artificial data.

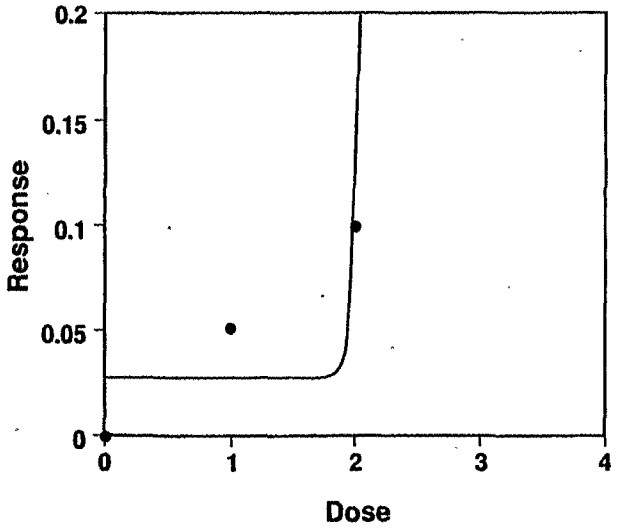

Figure 5c. "Implausible" fit of restricted Weibull model to artificial data.